

# Estadística Descriptiva

- **Población:** Es un conjunto de elementos con una determinada característica.
- **Muestra:** Es un subconjunto de la población.
- **Muestreo:** Es el proceso para elegir una muestra que sea representativa de la población.
- **Carácter estadístico:** Es la propiedad de los individuos que permite clasificarlos.
  - **Cualitativos:** No se pueden medir ni contar (color de ojos, periódico que lees).
  - **Cuantitativos:** Se pueden medir y contar (peso, talla, número de hermanos).

## Ejercicio 1.

Imaginemos las puntuaciones, entre 1 y 5, dadas por 50 alumnos referentes a su grado de satisfacción después de ver la última película de Indiana Jones y el Reino de la Calavera de Cristal.

3	5	3	4	2	3	1	3	4	2
3	4	5	3	1	2	4	3	1	5
3	2	2	4	4	5	1	2	4	1
1	2	3	4	2	5	4	4	1	2
5	3	3	4	1	3	3	2	4	3

- 1.- ¿Cuál ha sido la puntuación media?
- 2.- ¿Cuál es la puntuación más frecuente?
- 3.- ¿Cuántos alumnos han puntuado por debajo de 3?
- 4.- ¿Qué porcentaje de alumnos ha puntuado un 2?
- 5.- ¿Qué porcentaje de alumnos ha puntuado con un 4 o menos?

### Debemos organizar los datos:

- Primero definimos una variable X, cuyos valores sean:  
 $\{x_1, x_2, x_3, x_4, x_5\}$

$$x_1 = 1, \quad x_2 = 2, \quad x_3 = 3, \quad x_4 = 4, \quad x_5 = 5$$

- Llamamos, frecuencia absoluta,  $f_i =$  “número de veces que está el dato  $x_i$ ”.

- Ahora hacemos un recuento:

X	Recuento	f
$x_1 = 1$	//// /	8 = $f_1$
$x_2 = 2$	//// ////	10 = $f_2$
$x_3 = 3$	//// //// ///	14 = $f_3$
$x_4 = 4$	//// //// //	12 = $f_4$
$x_5 = 5$	//// /	6 = $f_5$

¿Cuál ha sido la puntuación media?

$$\frac{\text{suma de puntuaciones}}{50} = \frac{1 \cdot 8 + 2 \cdot 10 + 3 \cdot 14 + 4 \cdot 12 + 5 \cdot 6}{50} = \frac{148}{50} = 2'96$$

¿Cuál es la puntuación más frecuente?

La puntuación más frecuente ha sido el 3.

Debemos seguir organizando los datos:

Llamamos, frecuencia absoluta acumulada,  $F_i$  = “número de veces que se ha puntuado un número menor o igual que i”.

X	f	F
$x_1 = 1$	8 = $f_1$	8 = $F_1$
$x_2 = 2$	10 = $f_2$	8+10=18 = $F_2$
$x_3 = 3$	14 = $f_3$	8+10+14=32 = $F_3$
$x_4 = 4$	12 = $f_4$	8+10+14+12=44 = $F_4$
$x_5 = 5$	6 = $f_5$	8+10+14+12+6=50 = $F_5$

¿Cuántos alumnos han puntuado por debajo de 3?

Mirando la tabla tenemos que la respuesta es 18 alumnos.

Seguimos organizando los datos:

Definimos la columna **p**, porcentaje, como el cociente  $\frac{f}{50}$ , así en cada cuadro tendremos  $p_1, p_2, p_3, p_4$  y  $p_5$ . (También se le llama h=frecuencia relativa).

Definimos la columna **P**, porcentaje acumulado, como el cociente  $\frac{F}{50}$ , así en cada cuadro tendremos  $P_1, P_2, P_3, P_4$  y  $P_5$ . (También se le llama  $h$ =frecuencia relativa acumulada).

<b>X</b>	<b>f</b>	<b>p</b>	<b>F</b>	<b>P</b>
$x_1 = 1$	8	$8/50=0.16$	8	0.16
$x_2 = 2$	10	$10/50=0.20$	18	0.36
$x_3 = 3$	14	$14/50=0.28$	32	0.64
$x_4 = 4$	12	$12/50=0.24$	44	0.88
$x_5 = 5$	6	$6/50=0.12$	50	1.00

¿Qué porcentaje de alumnos ha puntuado un 2?

La respuesta es el 20 %.

¿Qué porcentaje de alumnos ha puntuado con un 4 o menos?

La respuesta es el 88 %.

## Ejercicio 2.

Se ha medido la estatura (talla) de 100 alumnos de 4º de ESO de un Instituto, obteniendo los siguientes resultados en centímetros:

160	156	173	173	184	183	169	155	160	157
172	168	156	174	160	167	185	161	175	167
158	161	167	153	159	147	172	173	161	162
169	177	145	150	162	159	170	185	158	168
183	164	169	153	155	174	153	164	180	181
157	179	166	160	178	166	175	180	168	160
175	151	170	161	148	170	167	150	170	175
148	166	159	176	170	156	159	168	152	178
154	163	175	180	165	169	155	163	177	160
176	172	175	148	154	177	158	170	182	178

Debemos organizar los datos:

Como la talla mínima es 145 y la máxima es 185 cm., agrupamos los datos en intervalos, por ejemplo así: [140,150), [150,160), [160,170), [170,180) y [180,190).

Hacemos el recuento:

Intervalos	Recuento	f
[140,150)	////	5
[150,160)	//// //// //// //// ////	24
[160,170)	//// //// //// //// //// //// //	32
[170,180)	//// //// //// //// //// //// //	29
[180,190)	//// ////	10
	<b>Tamaño de la muestra</b>	<b>100</b>

Así podemos afirmar que hay 24 alumnos cuya altura se encuentra entre 150 y 160 centímetros, o de otra forma “hay 24 alumnos cuya talla está alrededor de 155 centímetros”.  $155 = (150+160)/2$  es la **marca de clase** que aquí hará las funciones de  $x_2$ .

Intervalos	Marcas de clase=X	f
[140,150)	$x_1 = 145$	5
[150,160)	$x_2 = 155$	24
[160,170)	$x_3 = 165$	32
[170,180)	$x_4 = 175$	29
[180,190)	$x_5 = 185$	10
	<b>Tamaño de la muestra</b>	<b>100</b>

Y completando la tabla tengo:

Intervalos	Marcas de clase=X	f	p	F	P
[140,150)	145	5	0,05	5	0,05
[150,160)	155	24	0,24	29	0,29
[160,170)	165	32	0,32	61	0,61
[170,180)	175	29	0,29	90	0,90
[180,190)	185	10	0,10	100	1,00
	<b>Tamaño de la muestra</b>	<b>100</b>			

Ahora puedo responder a preguntas como:

- 1.- ¿Cuál es la talla media?
- 2.- ¿Cuál es la talla más frecuente?
- 3.- ¿Cuántos alumnos tienen una talla por debajo de 160 cm.?
- 4.- ¿Qué porcentaje de alumnos tienen la talla entorno a 185 cm.?
- 5.- ¿Qué porcentaje de alumnos tienen una talla menor o igual que 165 cm.?

## Medidas o parámetros

**Medidas de centralización:** Nos dan una idea del valor central o medio de los datos, alrededor del cuál se encuentran todos los datos.

**Media aritmética:**

$$\bar{X} = \frac{X_1 f_1 + X_2 f_2 + \dots + X_n f_n}{f_1 + f_2 + \dots + f_n} = \frac{\sum_{i=1}^n X_i f_i}{\sum_{i=1}^n f_i} = \frac{\sum_{i=1}^n X_i f_i}{N}$$

**Moda:** Mo, es el valor de los datos,  $x_i$ , que mayor frecuencia absoluta tiene.

**Mediana:** Me, es el valor de los datos, que se encuentra en medio, es decir, que la mitad de la muestra es menor que él y la otra mitad mayor.

Ejemplo: Se ordenan los datos crecientemente: **Me = 5**

1 1 2 2 2 2 3 3 4 4 4 **5** 5 5 6 6 7 7 7 7 8 9 9

O en una tabla de distribución de frecuencias:  $N = 23$ ,  $23/2 = 11'5$ ,  $Me = 5$

X	f	F
1	2	2
2	4	6
3	2	8
4	3	11
<b>5</b>	3	<b>14</b>
6	2	16
7	4	20
8	1	21
9	2	23

Ejemplo: Se ordenan los datos crecientemente: **Me = (4+5)/2 = 4'5**

1 1 2 2 2 2 3 3 4 4 **4** **5** 5 5 6 6 7 7 7 8 9 9

O en una tabla de distribución de frecuencias:  $N = 22$ ,  $22/2 = 11$ ,  $Me = (4+5)/2 = 4'5$

X	f	F
1	2	2
2	4	6
3	2	8
<b>4</b>	3	<b>11</b>
<b>5</b>	3	<b>14</b>
6	2	16
7	3	19
8	1	20
9	2	22

Con datos agrupados en intervalos funciona igual con la marcas de clase.

**Cuartiles:** Son similares a la Me. Dividen a los datos en 4 trozos con el 25% (o la cuarta parte) cada uno:

Ejemplo: Se ordenan los datos crecientemente:

1 1 2 2 2 2 3 3 4 4 4 5 5 5 6 6 7 7 7 8 9 9  
 $\underline{Q_1}$   $Me=Q_2$   $\underline{Q_3}$

$$Q_1 = 2, \quad Q_2 = Me = 4'5, \quad Q_3 = 7$$

O en una tabla de distribución de frecuencias:

$$(1/4)22 = 5'5 \quad (2/4)22 = 11 \quad (3/4)22 = 16'5$$

X	f	F
1	2	2
<b>2</b>	4	<b>6</b>
3	2	8
<b>4</b>	3	<b>11</b>
<b>5</b>	3	<b>14</b>
6	2	16
<b>7</b>	3	<b>19</b>
8	1	20
9	2	22

**Medidas de dispersión:** Me dan una idea de lo alejados o cercanos que están los datos respecto a un valor central como la media aritmética.

**Rango:** Es la diferencia entre el mayor y el menor valor de la variable.  
 En el ejemplo anterior:  $9 - 1 = 8$

**Desviación Media:**  $DM = \frac{\sum_{i=1}^n |x_i - \bar{x}| f_i}{N}$  “Es la media aritmética de los valores absolutos de las desviaciones de los datos respecto de la media aritmética”.

**Varianza:**  $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot f_i}{N} = \frac{\sum_{i=1}^n x_i^2 \cdot f_i}{N}$  “Es la media aritmética de los cuadrados de las desviaciones de los datos respecto de la media aritmética”

**Desviación típica o standar:**  $s = +\sqrt{s^2}$

**Ejemplo útil para calcular los parámetros:**

x	f	xf	/x-media/	/x-media/f	(x-media)^2	(x-media)^2*f
1	2	2	3,636	7,273	13,223	26,446
2	4	8	2,636	10,545	6,950	27,802
3	2	6	1,636	3,273	2,678	5,355
4	3	12	0,636	1,909	0,405	1,215
5	3	15	0,364	1,091	0,132	0,397
6	2	12	1,364	2,727	1,860	3,719
7	3	21	2,364	7,091	5,587	16,760
8	1	8	3,364	3,364	11,314	11,314
9	2	18	4,364	8,727	19,041	38,083
	<b>Suma:22</b>	<b>Suma:102</b>		<b>Suma:46,000</b>		<b>Suma:131,091</b>

Así obtendrías:

$$\bar{x} = \frac{102}{22} = 4'636 \quad DM = \frac{46}{22} = 2'091 \quad s^2 = \frac{131'091}{22} = 5'959 \quad s = \sqrt{5'959} = 2'441$$

Pero lo mejor que puedes hacer es aprender a **USAR LA CALCULADORA**.

**Ejercicio:** Con los resultados de la distribución de frecuencias de las tallas de 50 alumnos, calculamos los parámetros.

Intervalos	Marcas de clase=X	f	xf	/x-media/	/x-media/f	(x-media)^2	(x-media)^2*f
[140,150)	145	5	725	21,5	107,5	462,25	2311,25
[150,160)	155	24	3720	11,5	276	132,25	3174
[160,170)	165	32	5280	1,5	48	2,25	72
[170,180)	175	29	5075	8,5	246,5	72,25	2095,25
[180,190)	185	10	1850	18,5	185	342,25	3422,5
		<b>100</b>	<b>16650</b>		<b>863</b>		<b>11075</b>

$$\bar{x} = \frac{16650}{100} = 166'5 \quad DM = \frac{863}{100} = 8'63 \quad s^2 = \frac{11075}{100} = 110'75 \quad s = \sqrt{110'75} = 10'52$$

Pero lo mejor que puedes hacer es aprender a **USAR LA CALCULADORA**.

## Gráficos

Ahora pasamos a representar gráficamente los resultados obtenidos:

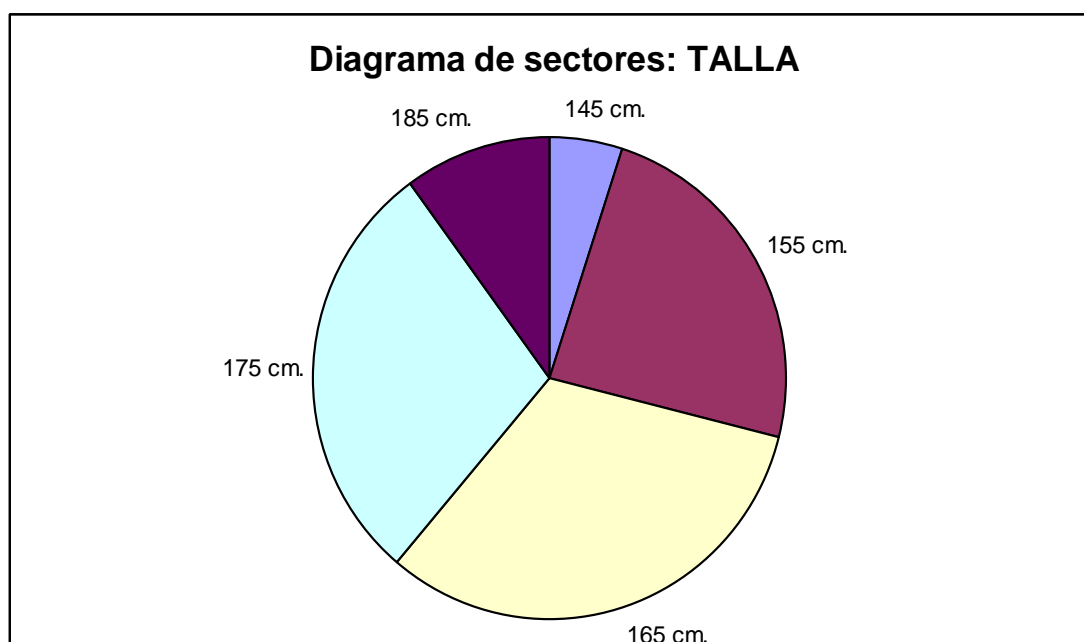
**Diagrama de sectores:** Con el ejemplo anterior, se trata de hacer pequeñas reglas de 3, si a los 100 alumnos les corresponden  $360^\circ$  (circunferencia entera), al dato  $x_i$  le corresponderán  $n_i^\circ$ .

Intervalos	Marcas de clase= $X$	f
[140,150)	145	5
[150,160)	155	24
[160,170)	165	32
[170,180)	175	29
[180,190)	185	10
		100

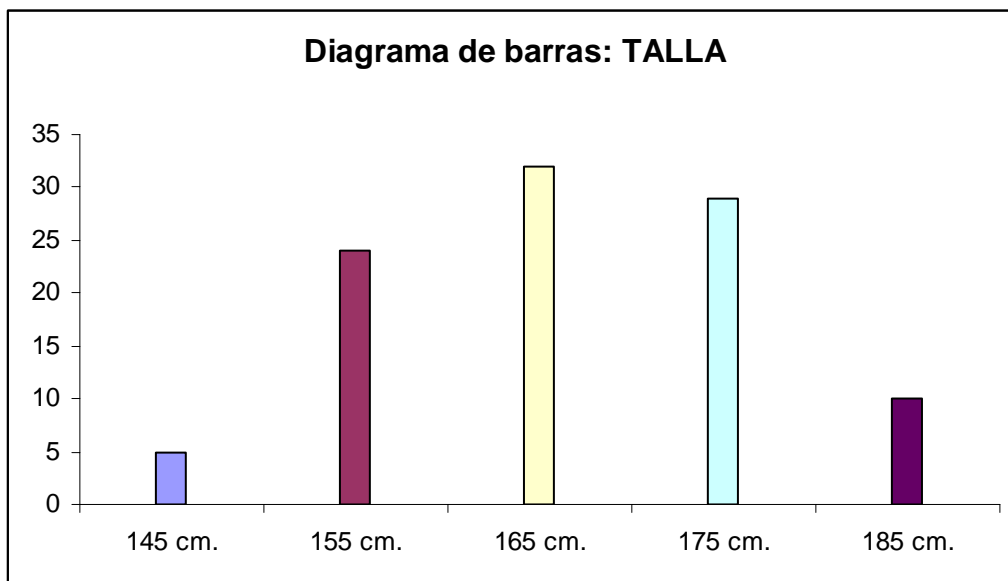
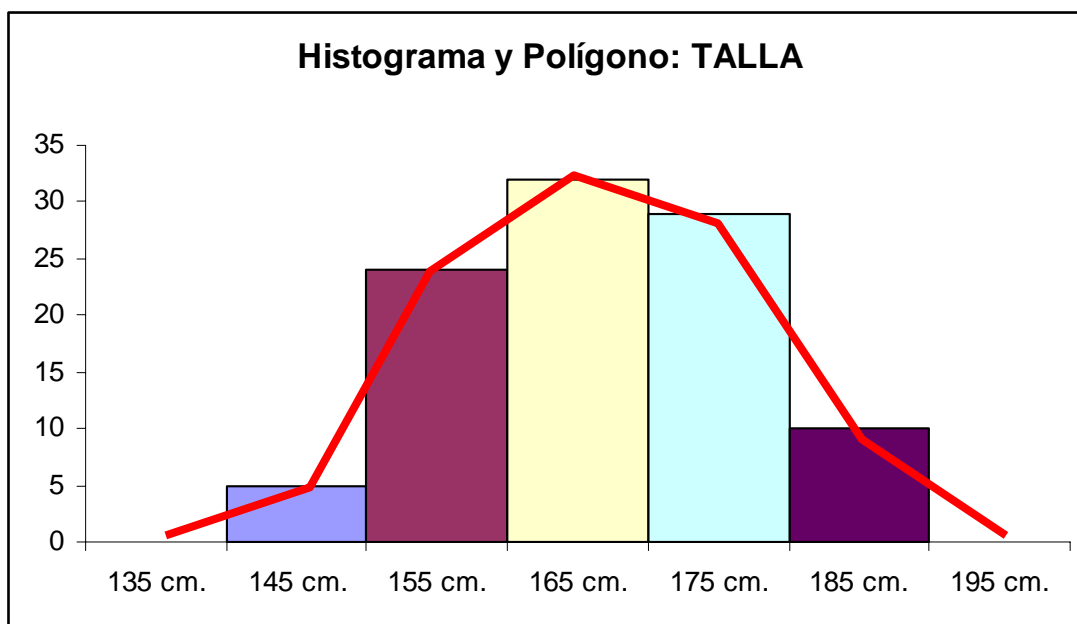
$$\left. \begin{array}{l} N = 100 \rightarrow 360^\circ \\ f_1 = 5 \rightarrow n_1^\circ \end{array} \right\} \Rightarrow n_1^\circ = \frac{5 \cdot 360^\circ}{100} = 5 \cdot \frac{360^\circ}{100} = 5 \cdot 3'6 = 18^\circ$$

Análogamente :

$$\begin{array}{l} \dots\dots\dots n_2^\circ = 24 \cdot 3'6 = 86'4^\circ \\ \dots\dots\dots n_3^\circ = 32 \cdot 3'6 = 115'2^\circ \\ \dots\dots\dots n_4^\circ = 29 \cdot 3'6 = 104'4^\circ \\ \dots\dots\dots n_5^\circ = 10 \cdot 3'6 = 36^\circ \\ \text{SUMA} = 360^\circ \end{array}$$





**Diagrama de barras:****Histograma y polígono de frecuencias:**

**Ejercicios:** Realiza idéntico trabajo con los siguientes datos recogidos:

1. Variable estadística: “Número de hermanos” N=50

2	3	1	1	1
0	2	5	4	1
1	0	2	0	2
3	2	1	2	1
1	1	3	1	4
0	0	1	1	2
3	3	0	2	2
4	1	2	0	1
0	4	2	0	0
1	0	1	3	1

2. Variable estadística: “Peso, en kg., de 80 personas” N=80

46,000	52,500	60,700	63,600	66,900	69,500	71,400	77,700
46,500	53,400	61,600	64,000	67,100	69,800	72,200	78,000
47,700	54,000	61,600	64,100	67,100	70,100	72,500	79,300
48,000	54,900	61,600	64,300	67,400	70,100	73,100	79,800
48,900	55,200	61,900	65,000	67,700	70,100	73,100	81,000
49,800	56,600	62,200	65,900	67,700	70,400	74,300	82,400
50,700	57,700	62,500	66,200	68,000	70,400	74,400	83,500
51,000	58,100	62,500	66,500	68,300	70,700	75,400	84,400
51,900	59,200	62,700	66,800	68,600	71,000	76,500	85,600
52,500	60,100	62,800	66,800	69,200	71,000	76,800	86,000

*Idea para confeccionar los intervalos:*

- 1.- El rango es  $86,000 - 46,000 = 40$ .
- 2.- Si deseamos, por ejemplo, hacer 7 intervalos (lo normal entre 6 y 15), pensamos en un número que añadir a 40 para que sea múltiplo de 7. Sería el 2.
- 3.- Así:  $(40+2)/7 = 6$  será la amplitud de cada intervalo.
- 4.- El 2 que hemos añadido lo repartimos entre el primer y último, así los intervalos serían:  $[45,000 - 51,000)$ ,  $[51,000 - 57,000)$ .....  $[81,000 - 87,000)$